

1 **Fitness tracking reveals task-specific associations**
2 **between memory, mental health, and exercise**

3 Jeremy R. Manning^{1,*}, Gina M. Notaro^{1,2}, Esme Chen¹, and Paxton C. Fitzpatrick¹

4 ¹Dartmouth College, Hanover, NH

5 ²Lockheed Martin, Bethesda, MD

6 *Address correspondence to jeremy.r.manning@dartmouth.edu

7 October 22, 2021

8 **Abstract**

9 Physical exercise can benefit both physical and mental well-being. Different forms of exer-
10 cise (i.e., aerobic versus anaerobic; running versus walking, swimming, or yoga; high-intensity
11 interval training versus endurance workouts; etc.) impact physical fitness in different ways. For
12 example, running may substantially impact leg and heart strength but only moderately impact
13 arm strength. We hypothesized that the mental benefits of exercise might be similarly differenti-
14 ated. We focused specifically on how different forms of exercise might relate to different aspects
15 of memory and mental health. To test our hypothesis, we collected (in aggregate) roughly a
16 century's worth of fitness data. We then asked participants to fill out surveys asking them to
17 self-report on different aspects of their mental health. We also asked participants to engage in
18 a battery of memory tasks that tested their short and long term episodic, semantic, and spatial
19 memory performance. We found that participants with similar exercise habits and fitness pro-
20 files tended to also exhibit similar mental health and task performance profiles. These effects
21 were task-specific in that different exercise patterns or fitness characteristics varied with different
22 aspects of memory, on different tasks.

²³ **Introduction**

²⁴ Engaging in physical activity (exercise) can improve physical fitness by increasing muscle strength
²⁵ (Crane et al., 2013; Knuttgen, 2007; Lindh, 1979; Rogers and Evans, 1993), bone density (Bassey and
²⁶ Ramsdale, 1994; Chilibeck et al., 2012; Layne and Nelson, 1999), cardiovascular performance (Maio-
²⁷ rana et al., 2000; Pollock et al., 2000), lung capacity (Lazovic-Popovic et al., 2016, although see Ro-
²⁸ man et al. (2016)), endurance (Wilmore and Knuttgen, 2003), and more. Exercise can also improve
²⁹ mental health (Basso and Suzuki, 2017; Callaghan, 2004; Deslandes et al., 2009; Mikkelsen et al.,
³⁰ 2017; Paluska and Schwenk, 2000; Raglin, 1990; Taylor et al., 1985) and cognitive performance (Basso
³¹ and Suzuki, 2017; Brisswalter et al., 2002; Chang et al., 2012; Etnier et al., 2006).

³² The physical benefits of exercise can be explained by stress-responses of the affected body tis-
³³ sues. For example, skeletal muscles that are taxed during exercise exhibit stress responses (Morton
³⁴ et al., 2009) that can in turn affect their growth or atrophy (Schiaffino et al., 2013). By comparison,
³⁵ the benefits of exercise on mental health are less direct. For example, one hypothesis is that ex-
³⁶ ercise leads to specific physiological changes, such as increased aminergic synaptic transmission
³⁷ and endorphin release, which in turn act on neurotransmitters in the brain (Paluska and Schwenk,
³⁸ 2000).

³⁹ Speculatively, if different exercise regimens lead to different neurophysiological responses, one
⁴⁰ might be able to map out a spectrum of signalling and transduction pathways that are impacted
⁴¹ by a given type, duration, and intensity of exercise in each brain region. For example, prior work
⁴² has shown that exercise increases acetylcholine levels, starting in the vicinity of the exercised
⁴³ muscles (Shoemaker et al., 1997). Acetylcholine is thought to play an important role in memory
⁴⁴ formation (e.g., by modulating specific synaptic inputs from entorhinal cortex to the hippocampus,
⁴⁵ albeit in rodents; Palacios-Filardo et al., 2021). Given the central role that these medial temporal
⁴⁶ lobe structures play in memory, changes in acetylcholine might lead to specific changes in memory
⁴⁷ formation and retrieval.

⁴⁸ In the present study, we hypothesize that (a) different exercise regimens will have different,
⁴⁹ quantifiable impacts on cognitive performance and mental health, and that (b) these impacts will

50 be consistant across individuals. To this end, we collected a year of fitness tracking data from each
51 of 113 participants. We then asked each participant to fill out a brief survey in which they self-
52 evaluated and self-reported several aspects of their mental health. Finally, we ran each participant
53 through a battery of memory tasks, which we used to evaluate their memory performance along
54 several dimensions. We searched the data for potential associations between memory, mental
55 health, and exercise.

56 **Methods**

57 We ran an online experiment using the Amazon Mechanical Turk (MTurk) platform (Gureckis
58 et al., 2016). We collected data about each participant’s fitness and exercise habits, a variety of
59 self-reported measures concerning their mental health, and about their performance on a battery
60 of memory tasks.

61 **Experiment**

62 **Participants**

63 We recruited experimental participants by posting our experiment as a Human Intelligence Task
64 (HIT) on the MTurk platform. We limited participation to MTurk Workers who had been assigned
65 a “master worker” designation on the platform, given to workers who score highly across several
66 metrics on a large number of HITs, according to a proprietary algorithm managed by Amazon.
67 One criteria embedded into the algorithm is a requirement that master workers must maintain a
68 HIT acceptance rate of at least 95%. We further limited our participant pool to participants who
69 self-reported that they were fluent in English and regularly used a Fitbit fitness tracker device.
70 A total of 160 workers accepted our HIT in order to participate in our experiment. Of these,
71 we excluded all participants who failed to log into their Fitbit account (giving us access to their
72 anonymized fitness tracking data), encountered technical issues (e.g., by accessing the HIT using an
73 incompatible browser, device, or operating system), or who ended their participation prematurely,

74 before completing the full study. In all, 113 participants contributed usable data to the study.
75 For their participation, workers received a base payment of \$5 per hour (computed in 15
76 minute increments, rounded up to the nearest 15 minutes), plus an additional performance-based
77 bonus of up to \$5. Our recruitment procedure and study protocol were approved by Dartmouth's
78 Committee for the Protection of Human Subjects.

79 **Gender, age, and race.** Of the 113 participants who contributed usable data, 77 reported their
80 gender as female, 35 as male, and 1 chose not to report their gender. Participants ranged in age
81 from 19–68 years old (25th percentile: 28.25 years; 50th percentile: 32 years; 75th percentile: 38
82 years). Participants reported their race as White (90 participants), Black or African American (11
83 participants), Asian (7 participants), Other (4 participants), and American Indian or Alaska Native
84 (3 participants). One participant opted not to report their race.

85 **Languages.** All participants reported that they were fluent in either 1 and 2 languages (25th
86 percentile: 1; 50th percentile: 1; 75th percentile: 1), and that they were "familiar" with between 1
87 and 11 languages (25th percentile: 1; 50th percentile: 2; 75th percentile: 3).

88 **Reported medical conditions and medications.** Participants reported having and/or taking med-
89 ications pertaining to the following medical conditions: anxiety or depression (4 participants),
90 recent head injury (2 participants), high blood pressure (1 participant), bipolar disorder (1 partici-
91 pant), hypothyroidism (1 participant), and other unspecified conditions or medications (1 partici-
92 pant). Participants reported their current and typical stress levels on a Likert scale as very relaxed
93 (-2), a little relaxed (-1), neutral (0), a little stressed (1), or very stressed (2). The "current" stress
94 level reflected participants' stress at the time they participated in the experiment. Their responses
95 ranged from -2 to 2 (current stress: 25th percentile: -2; 50th percentile: -1; 75th percentile: 1; typical
96 stress: 25th percentile: 0; 50th percentile: 1; 75th percentile: 1). Participants also reported their
97 current level of alertness on a Likert scale as very sluggish (-2), a little sluggish (-1), neutral (0), a
98 little alert (1), or very alert (2). Their responses ranged from -2 to 2 (25th percentile: 0; 50th per-
99 centile: 1; 75th percentile: 2). Nearly all (111 out of 113) participants reported that they had normal

¹⁰⁰ color vision, and 15 participants reported uncorrected visual impairments (including dyslexia and
¹⁰¹ uncorrected near- or far-sightedness).

¹⁰² **Residence and level of education.** Participants reported their residence as being located in the
¹⁰³ suburbs (36 participants), a large city (30 participants), a small city (23 participants), rural (14 partic-
¹⁰⁴ ipants), or a small town (10 participants). Participants reported their level of education as follows:
¹⁰⁵ College graduate (42 participants), Master's degree (23 participants), Some college (21 partici-
¹⁰⁶ pants), High school graduate (9 participants), Associate's degree (8 participants), Other graduate
¹⁰⁷ or professional school (5 participants), Some graduate training (3 participants), or Doctorate (2
¹⁰⁸ participants).

¹⁰⁹ **Reported water and coffee intake.** Participants reported the number of 8 oz cups of water and
¹¹⁰ coffee they had consumed prior to accepting the HIT. Water consumption ranged from 0–6 cups
¹¹¹ (25th percentile: 1; 50th percentile: 3; 75th percentile: 4). Coffee consumption ranged from 0–4 cups
¹¹² (25th percentile: 0; 50th percentile: 1; 75th percentile: 2).

¹¹³ **Tasks**

¹¹⁴ Upon accepting the HIT posted on MTurk, each worker was directed to read and fill out a screening
¹¹⁵ and consent form, and to share access to their anonymized Fitbit data via their Fitbit account. After
¹¹⁶ consenting to participate in our study and successfully sharing their Fitbit data, participants filled
¹¹⁷ out a survey and then engaged in a series of memory tasks (Fig. 1). All stimuli and code for running
¹¹⁸ the full MTurk experiment may be found [here](#).

¹¹⁹ **Survey questions.** We collected the following demographic information from each participant:
¹²⁰ their birth year, gender, highest (academic) degree achieved, race, language fluency, and language
¹²¹ familiarity. We also collected information about participants' health and wellness, including about
¹²² their vision, alertness, stress, sleep, coffee and water consumption, location of their residence,
¹²³ activity typically required for their job, and exercise habits.

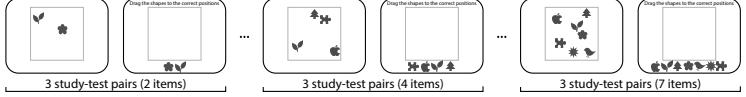
	Main task and immediate memory test				Delayed memory test
a.	1 Free recall	Study words 	Memory test 		5 
b.	2 Naturalistic recall	Watch a short video (The Temple of Knowledge) 	Memory tests Free response Multiple choice	6 	Free response
c.	3 Foreign language flashcards	Study flashcards 	Memory test Multiple choice	7 	Multiple choice
d.	4 Spatial learning	Memorize the positions of increasing numbers of shapes 			N/A

Figure 1: Battery of memory tasks. **a. Free recall.** Participants study 16 words (presented one at a time), followed by an immediate memory test where they type each word they remember from the just-studied list. In the delayed memory test, participants type any words they remember studying, from any list. **b. Naturalistic recall.** Participants watch a brief video, followed by two immediate memory tests. The first test asks participants to write out what happened in the video. The second test has participants answer a series of multiple choice questions about the conceptual content of the video. In the delayed memory test, participants (again) write out what happened in the video. **c. Foreign language flashcards.** Participants study a sequence of 10 English-Gaelic word pairs, each presented with an illustration of the given word. During an immediate memory test, participants perform a multiple choice test where they select the Gaelic word that corresponds to the given photograph. During the delayed memory test, participants perform a second multiple choice test, where they select the Gaelic word that corresponds to each of a new set of photographs. **d. Spatial learning.** In each trial, participants study a set of randomly positioned shapes. Next, the shapes' positions are altered, and participants are asked to drag the shapes back to their previous positions. **All panels.** The gray numbers denote the order in which participants experienced each task or test.

¹²⁴ **Free recall (Fig. 1a).** Participants studied a sequence of four word lists, each comprising 16 words.
¹²⁵ After studying each list, participants received an immediate memory test, whereby they were asked
¹²⁶ to type (one word at a time) any words they remembered from the just-studied list, in any order.

¹²⁷ Words were presented for 2 s each, in black text on a white background, followed by a 2 s blank
¹²⁸ (white) screen. After the final 2 s pause, participants were given 90 s to type in as many words
¹²⁹ as they could remember, in any order. The memory test was constructed such that the participant
¹³⁰ could only see the text of the current word they were typing; when they pressed any non-letter
¹³¹ key, the current word was submitted and the text box they were typing in was cleared. This was
¹³² intended to prevent participants from retroactively editing their previous responses.

¹³³ The word lists participants studied were drawn from the categorized lists reported by Ziman
¹³⁴ et al. (2018). Each participant was assigned four unique randomly chosen lists (in a randomized
¹³⁵ order), selected from a full set of 16 lists. Each chosen list was then randomly shuffled before
¹³⁶ presenting the words to the participants. Participants also performed a final delayed memory test
¹³⁷ where they were given 180 s to type out any words they remembered from *any* of the 4 lists they
¹³⁸ had studied.

¹³⁹ Recalled words within an edit distance of 2 (i.e., a Levenshtein Distance less than or equal to
¹⁴⁰ 2) of any word in the wordpool were “autocorrected” to their nearest match. We also manually
¹⁴¹ corrected clear typos or misspellings by hand (e.g., we corrected “hippopumas” to “hippopota-
¹⁴² mus”, “zucinni” to “zucchini”, and so on). Finally, we lemmatized each submitted word to match
¹⁴³ the plurality of the matching wordpool word (e.g., “bongo” was corrected to “bongos”, and so
¹⁴⁴ on). After applying these corrections, any submitted words that matched words presented on the
¹⁴⁵ just-studied list were tagged as “correct” recalls, and any non-matching words were discarded
¹⁴⁶ as “errors.” Because participants were not allowed to edit the text they entered, we chose not to
¹⁴⁷ analyze these putative “errors,” since we could not distinguish typos from true misrememberings.

¹⁴⁸ **Naturalistic recall (Fig. 1b).** Participants watched a 2.5 minute video clip entitled “The Temple
¹⁴⁹ of Knowledge.” The video comprises an animated story told to StoryCorps by Ronald Clark, who
¹⁵⁰ was interviewed by his daughter, Jamilah Clark. The narrator (Ronald) discusses growing up

¹⁵¹ living in an apartment over the Washington Heights branch of the New York Public Library, where
¹⁵² his father worked as a custodian during the 1940s.

¹⁵³ After watching the video clip, participants were asked to type out anything they remembered
¹⁵⁴ about what happened in the video. They typed their responses into a text box, one sentence at a
¹⁵⁵ time. When the participant pressed the return key or typed any final punctuation mark (".", "!", or
¹⁵⁶ "?") the text currently entered into the box was "submitted" and added to their transcript, and the
¹⁵⁷ text box was cleared to prevent further editing of any already-submitted text. This was intended to
¹⁵⁸ prevent participants from retroactively editing their previous responses. Participants were given
¹⁵⁹ up to 10 minutes to enter their responses. After 4 minutes, participants were given the option of
¹⁶⁰ ending the response period early, e.g., if they felt they had finished entering all of the information
¹⁶¹ they remembered. Each participant's transcript was constructed from their submitted responses by
¹⁶² combining the sentences into a single document and removing extraneous whitespace characters.
¹⁶³ Following this 4–10 minute free response period, participants were given a series of 10 multiple
¹⁶⁴ choice questions about the conceptual content of the story. All participants received the same
¹⁶⁵ questions, in the same order. Participants also performed a final delayed memory test, where they
¹⁶⁶ carried out the free response recall task a second time, near the end of the testing session. This
¹⁶⁷ resulted in a second transcript, for each participant.

¹⁶⁸ **Foreign language flashcards (Fig. 1c).** Participants studied a series of 10 English-Gaelic word
¹⁶⁹ pairs in a randomized order. We selected the Gaelic language both for its relatively small number
¹⁷⁰ of native speakers and for its dissimilarity to other commonly spoken languages amongst MTurk
¹⁷¹ workers. We verified (via self report) that all of our participants were fluent in English and that
¹⁷² they were neither fluent nor familiar with Gaelic.

¹⁷³ Each word's "flashcard" comprised a cartoon depicting the given word, the English word or
¹⁷⁴ phrase in lowercase text (e.g., "the boy"), and the Gaelic word or phrase in uppercase text (e.g.,
¹⁷⁵ "BUACHAILL"). Each flashcard was displayed for 4 s, followed by a 3 s interval (during which
¹⁷⁶ the screen was cleared) prior to the next flashcard presentation.

¹⁷⁷ After studying all 10 flashcards, participants were given a multiple choice memory test where

178 they were shown a series of novel photographs, each depicting one of the 10 words they had
179 learned. They were asked to select which (of 4 unique options) Gaelic word went with the given
180 picture. The 3 incorrect options were selected at random (with replacement across trials), and the
181 order in which the choices appeared to the participant were also randomized. Each of the 10 words
182 they had learned were tested exactly once.

183 Participants also performed a final delayed memory test, where they were given a second set of
184 10 questions (again, one per word they had studied). For this second set of questions participants
185 were prompted with a new set of novel photographs, and new randomly chosen incorrect choices
186 for each question. Each of the 10 original words they had learned were (again) tested exactly once
187 during this final memory test.

188 **Spatial learning (Fig. 1d).** Participants performed a series of study-test trials where they memo-
189 rized the onscreen spatial locations of two or more shapes. During the study phase of each trial,
190 a set of shapes appeared on the screen for 10 s, followed by 2 s of blank (white) screen. During the
191 test phase of each trial, the same shapes appeared onscreen again, but this time they were vertically
192 aligned and sorted horizontally in a random order. Participants were instructed to drag (using the
193 mouse) each shape to its studied position, and then to click a button to indicate that the placements
194 were complete.

195 In different study-test trials, participants learned the locations of different numbers of shapes
196 (always drawn from the same pool of 7 unique shapes, where each shape appeared at most one
197 time per trial). They first performed three trials where they learned the locations of 2 shapes; next
198 three trials where they learned the locations of 3 shapes; and so on until their last three trials, where
199 (during each trial) they learned the locations of 7 shapes. All told, each participant performed 18
200 study-test trials of this spatial learning task (3 trials for each of 2, 3, 4, 5, 6, and 7 shapes).

201 **Fitness tracking using Fitbit devices**

202 To gain access to our study, participants provided us with access to all data associated with their
203 Fitbit account from the year (365 calendar days) up to and including the day they accepted the HIT.

204 We filtered out all identifiable information (e.g., participant names, GPS coordinates, etc.) prior to
205 importing their data.

206 **Collecting and processing Fitbit data**

207 The fitness tracking data associated with participants' Fitbit accounts varied in scope and duration
208 according to which device the participant owned (Fig. S1), how often the participant wore (and/or
209 synced) their tracking device, and how long they had owned their device. For example, while all
210 participants' devices supported basic activity metrics such as daily step counts, only a subset of
211 the devices with heart rate monitoring capabilities provided information about workout intensity,
212 resting heart rate, and other related measures. Across all devices, we collected the following infor-
213 mation: heart rate data, sleep tracking data, logged bodyweight measurements, logged nutrition
214 measurements, Fitbit account and device settings, and activity metrics.

215 **Heart rate.** If available, we extracted all heart rate data collected by participants' Fitbit device(s)
216 and associated with their Fitbit profile. Depending on the specific device model(s) and settings, this
217 included second-by-second, minute-by-minute, daily summary, weekly summary, and/or monthly
218 summary heart rate information. These summaries include information about participants' aver-
219 age heart rates, and the amount of time they were estimated to have spent in different "heart rate
220 zones" (rest, out-of-range, fat burn, cardio, or peak, as defined by their Fitbit profile), as well as an
221 estimate of the number of estimated calories burned while in each heart rate zone.

222 **Sleep.** If available, we extracted all sleep data collected by participants' Fitbit device(s). Depend-
223 ing on the specific device model(s) and settings, this included nightly estimates of the duration
224 and quality of sleep, as well as the amount of time spent in each sleep stage (awake, REM, light, or
225 deep).

226 **Weight.** If available, we extracted any weight-related information affiliated with participants'
227 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific device
228 model(s) and settings, this included their weight, body mass index, and/or body fat percentage.

229 **Nutrition.** If available, we extracted any nutrition-related information affiliated with participants'
230 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific account
231 settings and usage behaviors, this included a log of the specific foods they had eaten (and logged)
232 over the past year, and the amount of water consumed (and logged) each day.

233 **Account and device settings.** We extracted any settings associated with participants' Fitbit ac-
234 counts to determine (a) which device(s) and model(s) are associated with their Fitbit account, (b)
235 time(s) when their device(s) were last synced, and (c) battery level(s).

236 **Activity metrics.** If available, we extracted any activity-related information affiliated with par-
237 ticipants' Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific
238 device model(s) and settings, this included: daily step counts; daily amount of time spent in each
239 activity level (sedentary, lightly active, fairly active, or very active, as defined by their account
240 settings and preferences); daily number of floors climbed; daily elevation change; and daily total
241 distance traveled.

242 **Comparing recent versus baseline measurements.**

243 We were interested in separating out potential associations between *absolute* fitness metrics and
244 *relative* metrics. To this end, in addition to assessing potential raw (absolute) fitness metrics, we
245 also defined a simple measure of recent changes in those metrics, relative to a baseline:

$$\Delta_{R,B}m = \frac{B \sum_{i=1}^R m(i)}{R \sum_{i=R+1}^{R+B} m(i)},$$

246 where $m(i)$ is the value of metric m from $i - 1$ days prior to testing (e.g., $m(1)$ represents the value
247 of m on the day the participant accepted the HIT, and $m(10)$ represents the value of m 9 days prior
248 to accepting the HIT. We set $R = 7$ and $B = 30$. In other words, to estimate recent changes in any
249 metric m , we divided the average value of m taken over the prior week by the average value of m
250 taken over the 30 days before that.

251 **Exploratory correlation analyses**

252 We used a bootstrap procedure to identify reliable correlations between different memory-related,
253 fitness-related, and demographic-related variables. For each of $N = 10,000$ iterations, we selected
254 (with replacement) a sample of 113 participants to include. This yielded, for each iteration, a
255 sampled “data matrix” with one row per sampled participant and one column for each measured
256 variable. When participants were sampled multiple times in a given iteration, as was often the
257 case, this matrix contained duplicate rows. Next, we computed the Pearson’s correlation between
258 each pair of columns. This yielded, for each pair of columns, a distribution of N bootstrapped
259 correlation coefficients. If 97.5% or fewer of the coefficients for a given pair of columns had the
260 same sign, we excluded the pair from further analysis and considered the expected correlation
261 between those columns to be undefined. If > 97.5% of the coefficients for a given pair of columns
262 had the same sign (corresponding to a bootstrap-estimated two-tailed p threshold of 0.05), we
263 computed the expected correlation coefficient as:

$$\mathbb{E}_{i,j} [r] = \tanh \left(\frac{1}{N} \sum_{n=1}^N \tanh^{-1} (\text{corr}(m(i)_n, m(j)_n)) \right),$$

264 where $m(x)_n$ represents column x of the bootstrapped data matrix for iteration n , \tanh is the
265 hyperbolic tangent, and \tanh^{-1} is the inverse hyperbolic tangent. We estimated the corresponding
266 p -values for these correlations as one minus the proportion of bootstrapped correlations with the
267 same sign, multiplied by two.

268 **Reverse correlation analyses**

269 We sought to characterize potential associations between the *dynamics* of participants’ fitness-
270 related activities leading up to the time they participated in a memory task and their performance
271 on the given task. For each fitness-related variable, we constructed a timeseries matrix whose rows
272 corresponded to timepoints (sampled once per day) leading up to the day the participant accepted
273 the HIT for our study, and whose columns corresponded to different participants. These matrices
274 often contained missing entries, since different participants’ Fitbit devices tracked fitness-related

275 activities differently. For example, participants whose Fitbit devices lacked heart rate sensors
276 would have missing entries for any heart rate-related variables. Or, if a given participant neglected
277 to wear their fitness tracker on a particular day, the column corresponding to that participant
278 would have missing entries for that day. To create stable estimates, we smoothed the timeseries of
279 each fitness measure using a sliding window of 1 week. In other words, for each fitness measure,
280 we replaced the “observed value” for each day with the average values of that measure (when
281 available) over the 7 day interval ending on the given day.

282 In addition to this set of matrices storing timeseries data for each fitness-related variable, we also
283 constructed a memory performance matrix, M , whose rows corresponded to different memory-
284 related variables, and whose columns corresponded to different participants. For example, one
285 row of the memory performance matrix reflected the average proportion of words (across lists)
286 that each participant remembered during the immediate free recall test, and so on.

287 Given a fitness timeseries matrix, F , we computed the weighted average and weighted standard
288 error of the mean of each row of F , where the weights were given by a particular memory-related
289 variable (row of M). For example, if F contained participants’ daily step counts, we could use
290 any row of M to compute a weighted average across any participants who contributed step count
291 data on each day. Choosing a row of M that corresponded to participants’ performance on the
292 naturalistic recall task would mean that participants who performed better on the naturalistic recall
293 task would contribute more to the weighted average timeseries of daily step counts. Specifically,
294 for each row, t , of F , we computed the weighted average (across the S participants) as:

$$\bar{f}(t) = \sum_{s=1}^S \hat{m}(s)F(t,s),$$

295 where \hat{m} denotes the normalized min-max scaling of m (the row of M corresponding to the chosen
296 memory-related variable):

$$\hat{m} = \frac{m}{\sum_{s=1}^S \hat{m}(s)},$$

297 where

$$\hat{m} = (1 - \epsilon) \frac{m - \min(m)}{\max(m) - \min(m)} + \epsilon.$$

298 Here, ϵ provides a lower bound on the influence of the lowest-weighted participant's data. We
299 defined $\epsilon = 0.001$, ensuring that the lowest-weighted participant had relatively low (but non-zero)
300 influence. We computed the weighted standard error of the mean as:

$$\text{SEM}_m(f(t)) = \frac{\left| \sum_{s=1}^S (F(t, s) - \bar{f}(t)) \right|}{\sqrt{S}}.$$

301 When a given row of F was missing data from one or more participants, those participants were
302 excluded from the weighted average for the corresponding timepoint and the weights (across all
303 remaining participants) were re-normalized to sum to 1. The above procedure yielded, for each
304 memory variable, a timeseries of weighted average (and weighted standard error of the mean)
305 fitness tracking values leading up to the day of the experiment.

306 Results

307 Before testing our main hypotheses, we examined the behavioral data from each of four memory
308 tasks: a random word list learning “free recall” task (Fig. 1); a naturalistic recall task whereby par-
309 ticipants watched a short video and then recounted the narrative; a foreign language “flashcards”
310 task; and a spatial learning task. Each of the first three tasks (free recall, naturalistic recall, and the
311 flashcards task) included both an immediate (short term) memory test and a delayed (long term)
312 memory test. The spatial learning task included only an immediate test. Participants in all four
313 tasks exhibited general trends and tendencies that have been previously reported in prior work.
314 We were also interested in characterizing the variability in task performance across participants.
315 For example, if all participants exhibited near-identical behaviors or performance on a given task,
316 we would be unable to identify how memory performance on that task varied with mental health
317 or exercise.

318 When participants engaged in free recall of random word lists, they displayed strong primacy

319 and recency effects (Murdock, 1962) on the immediate memory tests (as reflected by improved
320 memory for early and late list items; Fig. 2a, left and right panels). On the delayed memory test,
321 the recency effect was substantially diminished (Fig. 3a, left and right panels), consistent with
322 myriad previous studies (for review see Kahana, 2012). Participants also tended to cluster their
323 recalls according to the words' study positions (Kahana, 1996) on both the immediate (Fig. 2a,
324 middle panel) and delayed (Fig. 3a, middle panel) memory tests.

325 When participants engaged in naturalistic recall by recounting the narrative of a short story
326 video, they reliably and accurately remembered the major narrative events on both the immediate
327 (Fig. 2b) and delayed (Fig. 3b) tests. This is consistent with prior work showing that memory for
328 rich narratives is both detailed and accurate (Chen et al., 2017; Heusser et al., 2021).

329 Performance on the foreign language flashcards task (immediate: Fig. 2c; delayed: Fig. 3c)
330 varied substantially across participants, and did not show any clear serial position effects. Participants
331 also displayed substantial variation in performance on the spatial learning task (Fig. 2d).
332 In general, participants reported the shape's positions more accurately when there were fewer
333 shapes. However, both the baseline estimation accuracy and the rate of decrease in accuracy as a
334 function of increasing number of memorized locations varied substantially across participants.

335 In addition to observing substantial across-participant variability in memory performance,
336 we also observed substantial variability in participants' fitness and activity metrics (Fig. 4). We
337 examined recent measurements, averaged over the week prior to testing (Fig. 4a), baselined mea-
338 surements (average over the prior week, divided by the average over the preceding 30 days;
339 Fig. 4b), along with more gradually varying measures that tended to remain relatively static over
340 timescales of weeks to months (Fig. 4c). Figure S6 displays across-participant distributions for
341 a broad selection of these measures, and Figures S7, S8, S9, and S10 show different participants'
342 fitness metrics, broken down by their performance on different memory tasks.

343 We wondered about potential links between the different aspects of participants' data. For
344 example, if people who exercised in a particular way also tended to perform better on a given
345 memory task, this could suggest that either (a) some property intrinsic to participants who exer-
346 cised in a particular way might also affect their memory performance on the given task, and/or

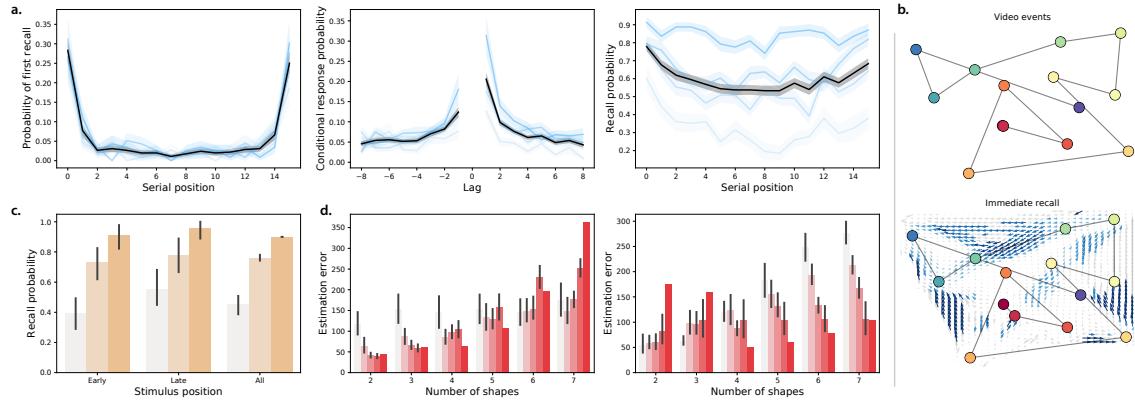


Figure 2: Immediate memory tests. **a. Free recall.** Left: probability of recalling each word first as a function of its presentation position. Middle: probability of transitioning between successively recalling the word presented at position i , followed by word presented at position $i + \text{Lag}$. Right: probability of recalling each word as a function of its presentation position. See Figure S2 for additional details. **b. Naturalistic recall.** Top: 2D embedding of a 2.5 min video clip; each dot reflects a narrative event (red denotes early events and blue denotes later events). Bottom: 2D embedding of the averaged transcripts of participants' recounts of the narrative (dots: same format as top panel). The arrows denote the average trajectory directions through the corresponding region of text embedding space, for any participants whose recounts passed through that region. Blue arrows denote statistically reliable agreement across participants ($p < 0.05$, corrected). See Figure S3 for additional details. **c. Foreign language flashcards.** Each bar denotes the average proportion of correctly recalled Gaelic-English word pairs from early (first 3), late (last 3), or all (i.e., all 10) study positions. See Figure S4 for additional details. **d. Spatial learning.** Average estimation error in shape locations as a function of the number of shapes. See Figure S5 for additional details. All panels: error bars and error ribbons denote bootstrap-estimated 95% confidence intervals. Shading (saturation) denotes results for different subsets of participants assigned based on their task performance (Figs. S2, S3, S4, and S5 provide information about which performance metrics and values the shading reflects; in general more saturated colors denote participants who performed better on the given task.) In Panel d, participants are grouped in two ways; in the left panel, participants are grouped according to the y -intercepts of regression lines (estimation error as a function of the number of shapes); in the right panel, participants are grouped according to the slopes of the same regression lines.

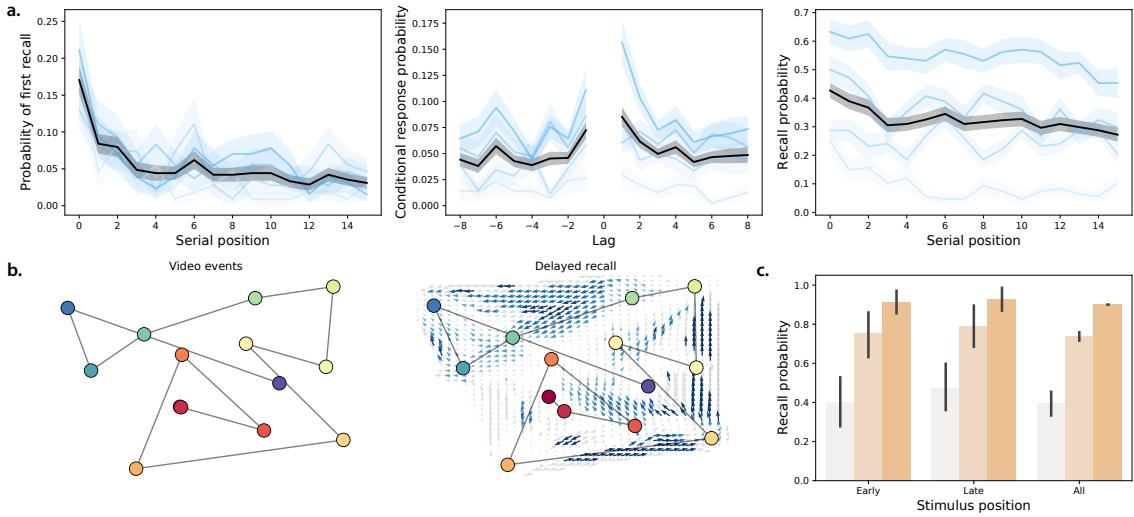


Figure 3: Delayed memory tests. **a. Free recall.** These panels are in the same format as Figure 2a, but they reflect performance on the delayed free recall task. For additional details see Figure S2. **b. Naturalistic recall.** These panels are in the same format as Figure 2b, but the right panel reflects performance on the delayed naturalistic recall task. For additional details see Figure S3. **c. Foreign language flashcards.** This panel is in the same format as Figure 2c, but it reflects performance on the delayed flashcards test. For additional details see Figure S4.

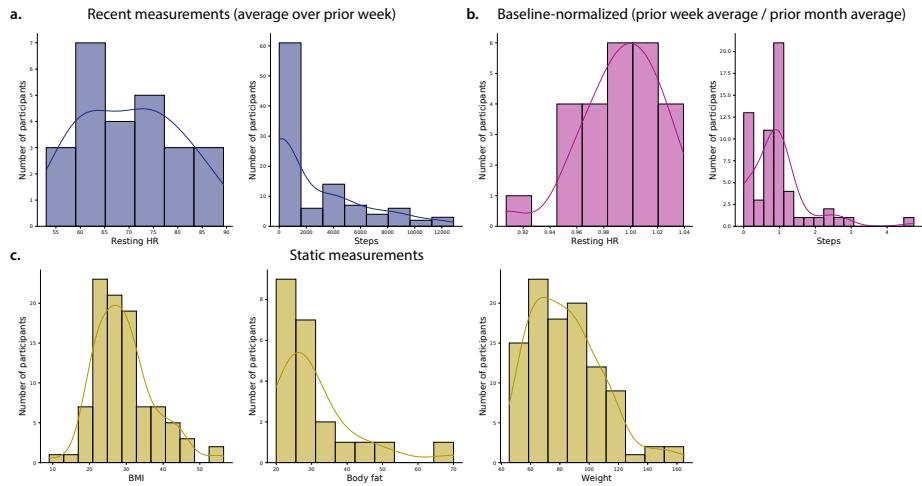


Figure 4: Fitness measures. **a. Recent measures.** Resting heart rate (HR) and daily step counts, averaged over the week prior to testing. **Baseline-normalized measures.** Resting heart rate and daily step counts averaged over the week prior to testing, divided by the average resting heart rate and step counts averaged over the preceding month. **Static measures.** Body mass index (BMI), body fat percentage, and weight (in kg). For more information see Figures S6, S7, S8, S9, and S10.

347 (b) particular exercise behaviors could have a causal impact on memory performance. We car-
348 ried out an exploratory analysis whereby we used a bootstrap-based approach (see *Exploratory*
349 *correlation analyses*) to identify reliable correlations between different aspects of memory perfor-
350 mance (Fig. S11), different aspects of fitness (Fig. S12), different demographic attributes (Fig. S13),
351 and correlations between memory performance, fitness information, and demographic attributes
352 (Fig. S14). Several patterns emerged. First, we found that participants' performance on the (within-
353 task) immediate versus delayed memory tests from the free recall, naturalistic recall, and foreign
354 language flashcards tasks were positively correlated ($rs > 0.25, ps < 0.003$). This suggests that,
355 within each of these tasks, similar processes or constraints may influence both short term and long
356 term information retrieval. We also found reliable across-task correlations between participants'
357 (immediate and delayed) performance on the free recall and foreign language flashcards tasks (rs
358 $> 0.3, ps < 0.03$).

359 A large number of fitness-related measures displayed reliable correlations (for a complete re-
360 port, see Fig. S12). For example, body mass index (BMI) and weight were correlated ($r = 0.91, p <$
361 0.0001). Resting heart rate over the prior week was negatively correlated with recent low-to-
362 moderate-intensity ("fat burn") cardiovascular activity levels ($r = 0.70, p = 0.0004$). Participants'
363 peak heart rates (averaged over the prior week) were also negatively correlated with recent in-
364 creases in step counts and daily elevation gains ($rs < -0.26, ps < 0.03$), where recent changes
365 were defined as the average values over the seven days leading up to the test day divided by
366 the average values over the preceding 30 days. Several demographic attributes (Fig. S13) dis-
367 played trivial correlations (e.g., participants identifying as male never reported identifying as
368 female, and so on). We also observed a negative correlation between reported stress and alertness
369 ($r = -0.44, p < 0.0001$), and positive correlations between the reported clarity of the instructions
370 for all tasks ($rs > 0.26, ps < 0.02$).

371 We also found reliable correlations between participants' fitness and demographic measures
372 and their behaviors in different tasks (for a complete report, see Fig. S14). For example, recent
373 low-to-moderate-intensity ("fat burn") cardiovascular activity was positively correlated with im-
374 mediate ($r = 0.38, p = 0.03$) and delayed ($r = 0.38, p = 0.029$) recall performance on the naturalistic

Behavioral measures	Mental health measures									
	Anxiety or depression	High blood pressure	Bipolar	Hypothyroid	Unspecified medications	Recent head injury	Current stress	Typical stress	Current / typical stress	Alertness
Free recall (immediate)		-0.11	0.11	-0.16	0.16	-0.05				
Free recall (delayed)			0.20	-0.16	0.10	-0.13		0.17		
Naturalistic recall (immediate)	0.07	0.12	-0.25			0.05				
Naturalistic recall (delayed)	-0.06	0.12	-0.15	0.10	-0.10					
Foreign language flashcards (immediate)	0.11				0.09	0.17				
Foreign language flashcards (delayed)		0.15		0.15			-0.29		-0.15	
Spatial learning (intercept)	0.02		-0.14	-0.10						
Spatial learning (slope)	0.04		-0.03	0.04	0.13				-0.19	

Figure 5: Memory performance differs according to mental health measures. The reported values in the table reflect correlations between each behavioral measure and mental health measure. Only statistically reliable correlations ($p < 0.05$, corrected) are displayed. We used participants' mean recall accuracy to characterize performance on the free recall and foreign language flashcards tasks, and mean precision to characterize performance on the naturalistic recall tasks. We characterized performance on the spatial learning task using the (inverted and normalized) intercepts and slopes of linear regressions on mean estimation errors as a function of the number of studied shapes (also see Figs. 2, 3, S2, S3, S4, and S5). Typical and current stress levels were measured by self report. Mental health information was inferred using each participants' list of self-reported medications.

375 memory task. Recent increases in moderate-intensity (“cardio”) activity over the prior 7 days
376 (relative to the preceding 30 days) was also positively correlated with immediate naturalistic re-
377 call performance ($r = 0.48, p = 0.003$) and immediate recall performance on the foreign language
378 flashcards task ($r = 0.43, p = 0.048$). Recent high-intensity (“peak”) activity was positively corre-
379 lated with performance on the spatial learning task ($r = 0.34, p < 0.0001$), as were recent increases
380 in high-intensity activity (prior 7 days versus preceding 30 days; $r = 0.41, p = 0.01$). Mental
381 health indicators, such as self-reported stress levels and medications were also associated with
382 differences in memory (Figs. 5, S14). For example, self-reported stress levels at the time of test
383 were negatively correlated with performance on the delayed memory test for the foreign language
384 flashcards task ($r = -0.29, p = 0.038$), whereas participants who were medicated for anxiety and
385 depression tended to perform slightly (but reliably) *better* on the immediate memory test for the
386 foreign language flashcards task ($r = 0.11, p < 0.0001$).

387 The above analyses indicate that recent differences in fitness-related activity are associated with
388 differences in memory performance and mental health measures. Although the analyses treated
389 these measures on average or in aggregate, many of the measures we collected are dynamic. For
390 example, the amount or intensity of physical exercise people engage in can vary over time, and
391 so on. We wondered whether the dynamics of fitness-related measures might relate to memory
392 performance and/or mental health measures. To this end, we carried out a series of reverse
393 correlation analyses (see *Reverse correlation analyses*) to examine whether participants with different
394 cognitive or mental health profiles also tended to display differences in fitness-related measures
395 over time. In particular, we examined fitness data collected from participants’ Fitbit devices over the
396 year prior to their test day in our study. Several example findings are summarized in Figure 6. We
397 found that participants who performed well on the immediate and delayed free recall memory tests
398 and on the naturalistic recall tests tended to be more active than participants who performed poorly
399 on those tests (Figs. 6a, b; S15). Conversely, participants who performed well on the immediate
400 and delayed foreign language flashcards tasks tended to be *less* active. These differences were
401 present even a full year before the testing day. We also found substantial variability across people
402 with different (self-reported) mental health profiles (Figs. 6c, S18). Due to small sample sizes of

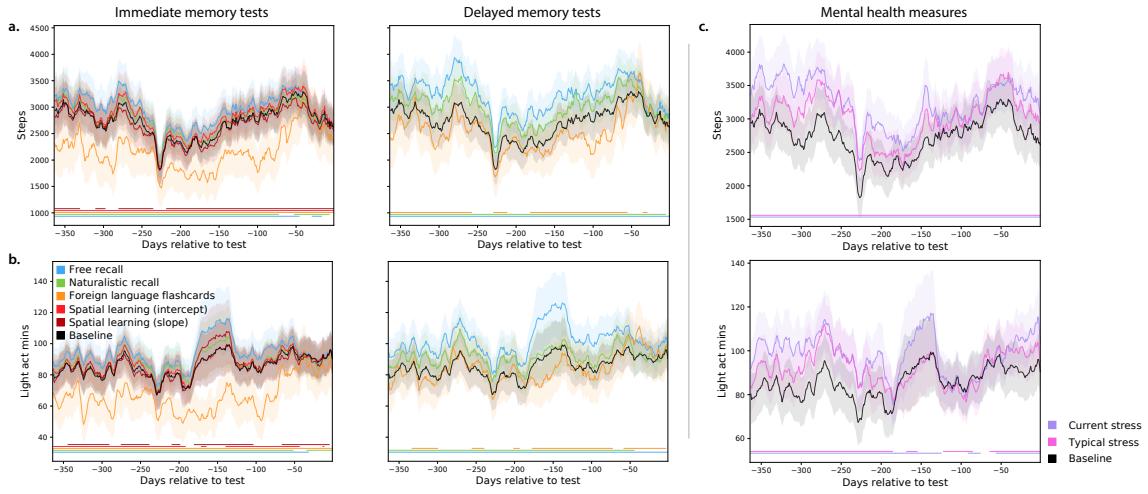


Figure 6: Dynamics of physical activity varies with memory performance and mental health measures. **a. Daily step counts.** Each timecourse is weighted by either performance on immediate recall tests (left panel) or on delayed recall tests (right panel). The black (baseline) timecourses display the (unweighted) average across all participants. **b. Daily duration (in minutes) of low-intensity physical activity.** Timecourses are displayed in the same format and color scheme as those in Panel A. Analogous timecourses for additional fitness-related measures may be found in Figures S15, S16, and S17. **c. Timecourses of physical activity, weighted by mental health measures.** The timecourses in each panel display the average daily step counts (top panel) or duration of low-intensity activity (bottom panel). The colored lines show average activity dynamics weighted by self-reported stress levels at the start of the experiment (purple) and self-reported “typical” stress levels (pink). The baseline curves (black) display the average across all participants (re-plotted in Panel C to illustrate scale differences across panels). Timecourses for additional mental health-related and fitness-related measures may be found in Figures S18, S19, and S20. Error ribbons in all panels denote the standard error of the mean. Horizontal lines below each panel’s timecourses denote intervals over which each weighted measure (color) differs from the unweighted baseline (via a paired sample two-sided t -test of the weighted mean values for each measure within a 30 day window around each timepoint; horizontal lines denote $p < 0.05$, corrected).

403 individuals exhibiting several mental health dimensions, it is difficult to distinguish generalizable
404 trends from individual differences that one or two individuals happened to exhibit. However,
405 several large-sample-size trends emerged. For example, participants who reported higher levels
406 of stress also tended to be slightly more physically active than participants who reported lower
407 stress levels. We found analogous differences in other activity-related measures (Figs. S15 and S18),
408 cardiovascular measures (Figs. S16 and S19), and sleep-related measures (Figs. S17 and S20). Taken
409 together, the analyses suggest that cognitive and mental health differences are also associated with
410 differences in the dynamics of physical health measures.

411 Discussion

412 After collecting a year's worth of fitness-tracking data from each of 113 participants, we ran each
413 participant in a battery of memory tasks and had them fill out a series of demographic and mental
414 health-related questions. We found that the associations between fitness-related activities, memory
415 performance, and mental health were heterogeneous. Our results suggest that engagement in
416 particular physical activities (e.g., differing in time relative to the test day, intensity, duration, etc.)
417 is also reflected in participants' memory performance patterns across tasks and in participants'
418 mental health attributes.

419 One important limitation of our study is that we cannot distinguish correlations between
420 different measures from potential causal effects. For example, we cannot know (from our study)
421 whether engaging in particular forms of exercise *causes* changes in memory performance or mental
422 health, or whether (alternatively) people who tend to engage in similar forms of exercise also
423 happen to exhibit similar memory and/or mental health profiles. In other words, an overlapping
424 set of processes or person-specific attributes may lead someone to both form particular habits
425 around exercise and display high or low performance on a given memory test. We do not know
426 whether memory performance or aspects of mental health might be manipulated or influenced
427 by changing the patterns of physical activity someone engages in. For this reason, we have been
428 careful to frame our findings as correlations and associations, rather than to imply knowledge

429 about a causal direction of our findings.

430 Although the present study cannot reveal causal effects, a large prior literature provides some
431 insight into potential causal effects by examining the neural and cognitive effects of a variety of
432 exercise interventions (Chang et al., 2015; Imboden et al., 2019; Kamijo et al., 2007; Sinha et al., 2021;
433 Suwabe et al., 2017; Tomporowski, 2003; Vidoni et al., 2015). A limitation of that prior work is that
434 most of these studies examine how relatively short-term changes in exercise (e.g., on timescales of
435 hours to days or, rarely, weeks to months) affect a cognitive performance on single task or aspect
436 of mental health. The present study examines longer-term exercise (over a full year), and relates
437 long-term exercise history to performance on a variety of tasks and to a variety of mental health
438 dimensions.

439 To the extent that exercise *does* provide a non-invasive means of manipulating cognitive per-
440 formance and mental health, our work may have exciting implications for cognitive enhancement.
441 For example, one might imagine building a recommendation system that suggests a particular ex-
442 ercise regimen tailored to improve a specific aspect of an individual's cognitive performance (e.g.,
443 the efficacy of a student's study session for an upcoming exam) or mental health (e.g., reducing
444 symptoms of anxiety before an important meeting). Just as strength training may be customized
445 to target a specific muscle group, or to improve performance on a specific physical task, similar
446 principles might also be applied to target specific improvements in cognitive fitness and mental
447 health.

448 Acknowledgements

449 We acknowledge useful discussions with David Bucci, Emily Glasser, Andrew Heusser, Abigail
450 Bartolome, Lorie Loeb, Lucy Owen, and Kirsten Ziman. Our work was supported in part by
451 the Dartmouth Young Minds and Brains initiative. The content is solely the responsibility of the
452 authors and does not necessarily represent the official views of our supporting organizations. This
453 paper is dedicated to the memory of David Bucci, who helped to inspire the theoretical foundations
454 of this work. Dave served as a mentor and colleague on the project prior to his passing.

455 **Data and code availability**

456 All analysis code and data used in the present manuscript may be found [here](#).

457 **Author contributions**

458 Concept: J.R.M. Experiment implementation and data collection: G.M.N. Analyses: G.M.N., E.C.,
459 P.C.F., and J.R.M. Writing: J.R.M. with input from all authors.

460 **Competing interests**

461 The authors declare no competing interests.

462 **References**

463 Bassey, E. J. and Ramsdale, S. J. (1994). Increase in femoral bone density in young women following
464 high-impact exercise. *Osteoporosis International*, 4:72–75.

465 Basso, J. C. and Suzuki, W. A. (2017). The effects of acute exercise on mood, cognition, neurophys-
466 iology, and neurochemical pathways: a review. *Brain Plasticity*, 2(2):127–152.

467 Brisswalter, J., Collardeau, M., and René, A. (2002). Effects of acute physical exercise characteristics
468 on cognitive performance. *Sports Medicine*, 32:555–566.

469 Callaghan, P. (2004). Exercise: a neglected intervention in mental health care? *Psychiatric and*
470 *Mental Health Nursing*, 11(4):476–483.

471 Chang, Y.-K., Chu, C.-H., Wang, C.-C., Wang, Y.-C., Song, T.-F., Tsai, C.-L., and Etnier, J. L. (2015).
472 Dose-response relation between exercise duration and cognition. *Medicine and Science in Sports*
473 *and Exercise*, 47:159–165.

- 474 Chang, Y. K., Labban, J. D., Gapin, J. I., and Etnier, J. L. (2012). The effects of acute exercise on
475 cognitive performance: a meta-analysis. *Brain Research*, 1453:87–101.
- 476 Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared
477 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,
478 20(1):115.
- 479 Chilibeck, P. D., Sale, D. G., and Webber, C. E. (2012). Exercise and bone mineral density. *Sports
480 Medicine*, 19:103–122.
- 481 Crane, J. D., MacNeil, L. G., and Tarnopolsky, M. A. (2013). Long-term aerobic exercise is associated
482 with greater muscle strength throughout the life span. *The Journals of Gerontology: Series A*,
483 68(6):631–638.
- 484 Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F. A. M. S.,
485 Coutinho, E. S. F., and Laks, J. (2009). Exercise and mental health: many reasons to move.
486 *Neuropsychobiology*, 59:191–198.
- 487 Etnier, J. L., Nowell, P. M., Landers, D. M., and Sibley, B. A. (2006). A meta-regression to examine the
488 relationship between aerobic fitness and cognitive performance. *Brain Research: Brain Research
489 Reviews*, 52(1):119–130.
- 490 Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D.,
491 Hamrick, J. B., and Chan, P. (2016). psiTurk: an open-source framework for conducting replicable
492 behavioral experiments online. *Behavior Research Methods*, 48(3):829–842.
- 493 Heusser, A. C., Fitzpatrick, P. C., and Manning, J. R. (2021). Geometric models reveal behavioral
494 and neural signatures of transforming naturalistic experiences into episodic memories. *Nature
495 Human Behavior*, 5:905–919.
- 496 Imboden, C., Gerber, M., Beck, J., Eckert, A., Pühse, U., Holsboer-Trachsler, E., and Hatzinger, M.
497 (2019). Effects of aerobic exercise as add-on treatment for inpatients with moderate to severe
498 depression on depression severity, sleep, cognition, psychological well-being, and biomarkers:

- 499 study protocol, description of study population, and manipulation check. *Frontiers in Psychiatry*,
500 10(262):doi.org/10.3389/fpsyg.2019.00262.
- 501 Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, 24:103–109.
- 502 Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press, New York, NY.
- 503 Kamijo, K., Nishihira, Y., Higashiura, T., and Kuroiwa, K. (2007). The interactive effect of exercise
504 intensity and task difficulty on human cognitive processing. *International Journal of Psychophysiology*,
505 65(2):114–121.
- 506 Knuttgen, H. G. (2007). Strength training and aerobic exercise: comparison and contrast. *Journal of
507 Strength and Conditioning Research*, 21(3):973–978.
- 508 Layne, J. E. and Nelson, M. E. (1999). The effects of progressive resistance training on bone density:
509 a review. *Medicine and Science in Sports and Exercise*, 31(1):25–30.
- 510 Lazovic-Popovic, B., Zlatkovic-Svenda, M., Durmic, T., Djelic, M., Saranovic, D., and Zugic, V.
511 (2016). Superior lung capacity in swimmers: some questions, more answers! *Revista Portuguesa
512 de Pneumologia*, 22(3):151–156.
- 513 Lindh, M. (1979). Increase of muscle strength from isometric quadriceps exercises at different knee
514 angles. *Scandinavian Journal of Rehabilitation Medicine*, 11(1):33–36.
- 515 Maiorana, A., O'Driscoll, G., Cheetham, C., Collis, J., Goodman, C., Rankin, S., Taylor, R., and
516 Green, D. (2000). Combined aerobic and resistance exercise training improves functional capacity
517 and strength in CHF. *Journal of Applied Physiology*, 88(1565–1570).
- 518 Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M., and Apostolopoulos, V. (2017).
519 Exercise and mental health. *Maturitas*, 106:48–56.
- 520 Morton, J. P., Kayani, A. C., McArdle, A., and Drust, B. (2009). The exercise-induced stress response
521 of skeletal muscle, with specific emphasis on humans. *Sports Medicine*, 39:643–662.

- 522 Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology: General*, 64:482–488.
- 523
- 524 Palacios-Filardo, J., Udakis, M., Brown, G. A., Tehan, B. G., Congreve, M. S., Nathan, P. J., Brown, A. J. H., and Mellor, J. R. (2021). Acetylcholine prioritises direct synaptic inputs from entorhinal cortex to CA1 by differential modulation of feedforward inhibitory circuits. *Nature Communications*, 12(5475):doi.org/10.1038/s41467-021-25280-5.
- 525
- 526
- 527
- 528 Paluska, S. A. and Schwenk, T. L. (2000). Physical activity and mental health. *Sports Medicine*, 29(3):167–180.
- 529
- 530 Pollock, M. L., Franklin, B. A., Balady, G. J., Chaltman, B. L., Fleg, J. L., Fletcher, B., Limacher, M., na, I. L. P., Stein, R. A., Williams, M., and Bazzarre, T. (2000). Resistance exercise in individuals with and without cardiovascular disease. *Circulation*, 101:828–833.
- 531
- 532
- 533 Raglin, J. S. (1990). Exercise and mental health. *Sports Medicine*, 9:323–329.
- 534 Rogers, M. A. and Evans, W. J. (1993). Changes in skeletal muscle with aging: effects of exercise training. *Exercise and Sport Sciences Reviews*, 21:65–102.
- 535
- 536 Roman, M. A., Rossiter, H. B., and Casaburi, R. (2016). Exercise, ageing and the lung. *European Respiratory Journal*, 48:1471–1486.
- 537
- 538 Schiaffino, S., Dyar, K. A., Ciciliot, S., Blaauw, B., and Sandri, M. (2013). Mechanisms regulating skeletal muscle growth and atrophy. *The febs Journal*, 280(17):4294–4314.
- 539
- 540 Shoemaker, J. K., Halliwill, J. R., Hughson, R. L., and Joyner, M. J. (1997). Contributions of acetylcholine and nitric oxide to forearm blood flow at exercise onset and recovery. *Vascular Physiology*, 273(5):2388–2395.
- 541
- 542
- 543 Sinha, N., Berg, C. N., Yassa, M. A., and Gluck, M. A. (2021). Increased dynamic flexibility in the medial temporal lobe network following an exercise intervention mediates generalization of prior learning. *Neurobiology of Learning and Memory*, 177:107340.
- 544
- 545

- 546 Suwabe, K., Hyodo, K., Byun, K., Ochi, G., Yassa, M. A., and Soya, H. (2017). Acute moderate
547 exercise improves mnemonic discrimination in young adults. *Hippocampus*, 27(3):229–234.
- 548 Taylor, C. B., Sallis, J. F., and Needle, R. (1985). The relation of physical activity and exercise to
549 mental health. *Public Health Reports*, 100(2):195–202.
- 550 Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica*,
551 112(3):297–324.
- 552 Vidoni, E. D., Johnson, D. K., Morris, J. K., Van Sciver, A., Greer, C. S., Billinger, S. A., Donnelly, J. E.,
553 and Burns, J. M. (2015). Dose-response of aerobic exercise on cognition: A community-based,
554 pilot randomized controlled trial. *PLoS One*, 10(7):1–13.
- 555 Wilmore, J. H. and Knuttgen, H. G. (2003). Aerobic exercise and endurance. *The Physician and
556 Sportsmedicine*, 31(5):45–51.
- 557 Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., and Manning, J. R. (2018). Is automatic
558 speech-to-text transcription ready for use in psychological experiments? *Behavior Research
559 Methods*, 50:2597–2605.